#12
28
6/18/03

01263.001203

PATENT APPLICATION

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of:	)	
	:	Examiner: Donald L. Storm
JASON PETER ANDREW CHARLESWORTH, ET. AL.	)	
	:	
	)	Group Art Unit: 2654
Application No.: 09/934,799	:	
	)	
Filed: August 23, 2001	:	
	)	
For: DATABASE ANNOTATION AND RETRIEVAL INCLUDING PHONEME DATA (AS AMENDED)	:	June 12, 2003
	)	
	:	

Mail Stop Non-Fee Amendment
COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450

SUBMISSION OF PRIORITY DOCUMENTS

Sir:

In support of Applicants' claim for priority under 35 U.S.C. § 119, enclosed are certified copies of the following foreign applications:

United Kingdom	9905201.1	March 5, 1999;
United Kingdom	9905186.4	March 5, 1999;
United Kingdom	9905187.2	March 5, 1999;
United Kingdom	9905199.7	March 5, 1999; and
United Kingdom	9905160.9	March 5, 1999.

Applicants' undersigned attorney may be reached in our Washington, D.C. office by telephone at (202) 530-1010. All correspondence should continue to be directed to our address given below.

Respectfully submitted,

Shawn W. Fraser

Attorney for Applicants
Shawn W. Fraser
Registration No. 45,886

FITZPATRICK, CELLA, HARPER & SCINTO
30 Rockefeller Plaza
New York, New York 10112-3801
Facsimile: (212) 218-2200

SWF:eyw

DC_MAIN 135623 v 1

09/934,799 #12

INVESTOR IN PEOPLE

CERTIFIED COPY OF PRIORITY DOCUMENT

The Patent Office
Concept House
Cardiff Road
Newport
South Wales
NP10 8QQ

, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) f the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the omptroller-General, hereby certify that annexed hereto is a true copy of the documents as iginally filed in connection with the patent application identified therein.

accordance with the Patents (Companies Re-registration) Rules 1982, if a company named his certificate and any accompanying documents has re-registered under the Companies Act 0 with the same name as that with which it was registered immediately before re-stration save for the substitution as, or inclusion as, the last part of the name of the words blic limited company" or their equivalents in Welsh, references to the name of the company is certificate and any accompanying documents shall be treated as references to the name which it is so re-registered.

ordance with the rules, the words "public limited company" may be replaced by p.l.c., .L.C. or PLC.

istration under the Companies Act does not constitute a new legal entity but merely s the company to certain additional company law rules.

Signed

Dated 9 April 2003

Patents Form 1/77
Patents Act 1977
(Rule 16)

The Patent Office

Request for grant of a patent

The Patent Office
Cardiff Road
Newport
Gwent NP9 1RH

1. Your reference
2644601/AM

2. Patent Application Number
9905187.2

3. Full name, address and postcode of the or of each applicant *(underline all surnames)*

Canon Kabushiki Kaisha
30-2 3-Chome Shimomaruko
Ohta-Ku
Tokyo
Japan

Patents ADP number *(if known)* 6387658001

If the applicant is a corporate body, give the country/state of its incorporation — Country: JAPAN
State:

4. Title of the invention
DATABASE ANNOTATION AND RETRIEVAL

5. Name of agent — **Beresford & Co**

"Address for Service" in the United Kingdom to which all correspondence should be sent

2/5 Warwick Court
High Holborn
London WC1R 5DJ

Patents ADP number 1826001

6. Priority details

Country	Priority application number	Date of filing

Patents Form 1/77

7. If this application is divided or otherwise derived from an earlier UK application give details

Number of earlier of application | Date of filing

8. Is a statement of inventorship and or right to grant of a patent required in support of this request?

YES

9. Enter the number of sheets for any of the following items you are filing with this form.

Continuation sheets of this form	
Description	19
Claim(s)	5
Abstract	1
Drawing(s)	11

10. If you are also filing any of the following, state how many against each item.

Priority documents	
Translations of priority documents	
Statement of inventorship and right to grant of a patent *(Patents form 7/77)*	1 + 3 COPIES
Request for preliminary examination and search *(Patents Form 9/77)*	
Request for Substantive Examination *(Patents Form 10/77)*	
Any other documents *(please specify)*	

11. I/We request the grant of a patent on the basis of this application

Signature Beresford & Co Date 5 March 1999

BERESFORD & Co

12. Name and daytime telephone number of person to contact in the United Kingdom

ALAN MACDOUGALL

Tel:0171-831-2290

Patents Form 7/77
Patents Act 1977
(Rule 15)

The Patent Office

Statement of inventorship and of right to grant of a patent

The Patent Office
Cardiff Road
Newport
Gwent NP9 1RH

1. Your reference
2644601/AM

- 5 MAR 1999

2. Patent Application Number
accompanying application reference 2644601

9905187.2

3. Full name of the or each applicant

Canon Kabushiki Kaisha

4. Title of the invention
DATABASE ANNOTATION AND RETRIEVAL

5. State how the applicant(s) derived the right from the inventor(s) to be granted a patent

By employment of the inventors by Canon Research Centre Europe Limited, and by a general agreement dated 1 January 1994 between Canon Research Centre Europe Limited and the applicant.

6. How many, if any additional Patents Forms 7/77 are attached to this form?

NONE

11. I/We believe that the person(s) named over the page (and on any extra copies of this form) is/are the inventor(s) of the invention which the above patent application relates to.

Signature Beresford & Co Date 5 March 1999
BERESFORD & Co

12. Name and daytime telephone number of person to contact in the United Kingdom

ALAN MACDOUGALL

Tel: 0171-831-2290

Patents Form 7/77

CHARLESWORTH, Jason Peter Andrew
c/o Canon Research Centre Europe Limited
1 Occam Court, Occam Road
Surrey Research Park
Guildford
Surrey GU2 5YJ 7615701001

RAJAN, Jebu Jacob
c/o Canon Research Centre Europe Limited
1 Occam Court, Occam Road
Surrey Research Park
Guildford
Surrey GU2 5YJ 7615719001

GARNER, Philip Neil
c/o Canon Research Centre Europe Limited
1 Occam Court, Occam Road
Surrey Research Park
Guildford
Surrey GU2 5YJ 7615727001

DATABASE ANNOTATION AND RETRIEVAL

The present invention relates to the annotation of data files which are to be stored in a database for facilitating their subsequent retrieval. The present invention is also concerned with a system for generating the annotation data which is added to the data file and to a system for searching the annotation data in the database to retrieve a desired data file in response to a user's input query.

Databases of information are well known and suffer from the problem of how to locate and retrieve the desired information from the database quickly and efficiently. Existing database search tools allow the user to search the database using typed keywords. Whilst this is quick and efficient, this type of searching is not suitable for various kinds of databases, such as video or audio databases.

According to one aspect, the present invention aims to provide a data structure which will allow the annotation of data files within a database which will allow a quick and efficient search to be carried out in response to a user's input query.

Exemplary embodiments of the present invention will now

be described with reference to Figures 1 to 10, in which:

Figure 1 is a block schematic diagram showing the form of a document annotation system embodying the present invention;

Figure 2 is a schematic diagram of a word and phoneme lattice embodying one aspect of the present invention, for an example text string from an input document;

Figure 3 illustrates the form of a document annotation system embodying the present invention;

Figure 4 is a schematic block diagram of a document annotation system embodying the present invention.

Figure 5 is a schematic block diagram of a user's terminal which allows the user to retrieve information from the database by a voice query;

Figure 6a is a flow diagram illustrating part of the flow control of the user terminal shown in Figure 5;

Figure 6b is a flow diagram illustrating the remaining part of the flow control of the user terminal shown in Figure 5;

Figure 7 is a flow diagram illustrating the way in which a search engine forming part of the user's terminal carries out a phoneme search within the database;

Figure 8 is a schematic diagram illustrating the form of a phoneme string and four M-GRAMS generated from the phoneme string;

Figure 9 is a plot showing two vectors and the angle between the two vectors;

Figure 10 is a schematic block diagram illustrating a user terminal which is operable to access a database located on a remote server via a data network in response to an input utterance by the user; and

Figure 11 is a schematic block diagram of a user terminal which allows a user to access a database located in a remote server in response to an input utterance from the user.

Embodiments of the present invention can be implemented using dedicated hardware circuits, but the embodiment to be described is implemented in computer software or code, which is run in conjunction with processing hardware such as a personal computer, work station, photocopier, facsimile machine, personal digital assistant (PDA) or

the like.

DATA FILE ANNOTATION

Figure 1 is a block diagram illustrating the way in which annotation data is generated for and appended to a document before storage in a database 29. In particular, as shown in Figure 1 a text document 101 is converted into an image data file by a document scanner 103. The image data file is then passed to an optical character recognition (OCR) unit 105 which converts the image data of the document 101 into electronic text. This electronic text is then supplied to a phonetic transcription unit 107 which is operable to generate phoneme (or phoneme like) and word annotation data 109 which is then appended to the image data output by the scanner 103 to form a data file 111. As shown, the data file 111 is then stored in the database 29 for subsequent retrieval.

In this embodiment, the annotation data 109 comprises a combined phoneme and word lattice which allows the user to subsequently retrieve the data file 111 from the database 29 by a voice query. The use of both phoneme and word data allows a quick and efficient search of the database 29 to be carried out. In particular, a quick search using the word data can be carried out and, if this fails to provide the required data file 111, then

a further search using the more robust phoneme data can be performed. As those skilled in the art of speech recognition will realise, the use of phoneme data is more robust, because phonemes are dictionary independent and allow the system to cope with out of vocabulary words, such as names, places, foreign words etc. The use of phoneme data is also capable of making the system future proof, since it allows data files which are placed into the database to be retrieved even when the optical character recognition unit 105 makes a character recognition mistake in converting the image data into text data.

Figure 2 illustrates the form of the phoneme and word lattice annotation data output by the phonetic transcription unit 107 for the input text '...tell me about Jason...'. As shown, the phonetic transcription unit 107 identifies a number of different possible phoneme strings which correspond to this input text, using a text to phoneme dictionary (not shown). A number of different possible phoneme strings are generated in order to provide flexibility during the subsequent retrieval since the retrieval process may be carried out in response to an input voice query. As shown in Figure 2, the phoneme and word lattice is an acyclic directed graph with a single entry point and a single exit point. It represents different parses of the text stream output

by the optical character recognition unit 105.

In this embodiment, the annotation data 109 stored in the database 29 has the following general form:

HEADER

- time index associating the location of blocks of annotation data within memory to a given time point.
- word set used (i.e. the dictionary)
- phoneme set used
- the language to which the vocabulary pertains

Block(i) $i = 0,1,2,.....$

node N_j $j = 0,1,2,.....$

- time offset of node from start of block
- phoneme links (k) $k = 0,1,2.....$
  offset to node $N_j = N_k - N_j$ (N_k is node to which link K extends)
  phoneme associated with link (k)
- word links (l) $l = 0,1,2,.....$
  offset to node $N_j = N_i - N_j$ (N_j is node to which link l extends)
  word associated with link (l)

In this embodiment, the annotation data is divided into blocks of nodes in order to allow the search to jump into

the middle of the annotation data for a given audio data stream. The header therefore includes a time index which associates the location of the blocks of annotation data within the memory to a given time offset between the time of start and the time corresponding to the beginning of the block.

The header also includes data defining the word set used (i.e. the dictionary), the phoneme set used and the language to which the vocabulary pertains. The header may also include details of the automatic speech recognition system used to generate the annotation data and any appropriate settings thereof which were used during the generation of the annotation data.

The blocks of annotation data then follow the header and identify, for each node in the block, the time offset of the node from the start of the block, the phoneme links which connect that node to other nodes by phonemes and word links which connect that node to other nodes by words. Each phoneme link and word link identifies the phoneme or word which is associated with the link. They also identify the offset to the current node. For example, if node N_{50} is linked to node N_{55} by a phoneme link, then the offset to node N_{50} is 5. As those skilled in the art will appreciate, using an offset indication like this allows the division of the continuous

annotation data into separate blocks.

Figure 3 illustrates a modification to the document annotation system shown in Figure 1. The difference between the system shown in Figure 3 and the system shown in Figure 1 is that the output of the optical character recognition unit 105 is used to generate the data file 113, rather than the image data output by the scanner 103. The rest of the system shown in Figure 3 is the same as that shown in Figure 1 and will not be described further.

Figure 4 shows a further modification to the document annotation system shown in Figure 1. In the embodiment shown in Figure 4, the input document is received by a facsimile unit 115 rather than a scanner 103. The image data output by the facsimile unit is then processed in the same manner as the image data output by the scanner 103 shown in Figure 1, and will not be described again.

DATA FILE RETRIEVAL

Figure 5 is a block diagram illustrating the form of a user terminal 59 which can be used to retrieve the annotated data files from the database 29. This user terminal 59 may be, for example, a personal computer, a copying machine, a hand held device or the like. As shown, in this embodiment, the user terminal 59 comprises

the database 29 of annotated data files, an automatic speech recognition unit 51, a search engine 53, a control unit 55 and a display 57. In operation, the automatic speech recognition unit 51 is operable to process an input voice query from the user 39 received via the microphone 7 and the input line 61 and to generate therefrom corresponding phoneme and word data. This data may also take the form of a phoneme and word lattice, but this is not essential. This phoneme and word data is then input to the control unit 55 which is operable to initiate an appropriate search of the database 29 using the search engine 53. The results of the search, generated by the search engine 53, are then transmitted back to the control unit 55 which analyses the search results and generates and displays appropriate display data to the user via the display 57.

Figures 6a and 6b are flow diagrams which illustrate the way in which the user terminal 59 operates in this embodiment. In step s1, the user terminal 59 is in an idle state and awaits an input query from the user 39. Upon receipt of an input query, the phoneme and word data for the input query is generated in step s3 by the automatic speech recognition unit 51. The control unit 55 then instructs the search engine 53, in step s5, to perform a search in the database 29 using the word data generated for the input query. The word search employed

in this embodiment is the same as is currently being used in the art for typed keyword searches, and will not be described in more detail here. If in step s7, the control unit 55 identifies from the search results, that a match for the user's input query has been found, then it outputs the search results to the user via the display 57.

In this embodiment, the user terminal 59 then allows the user to consider the search results and awaits the user's confirmation as to whether or not the results correspond to the information the user requires. If they are, then the processing proceeds from step s11 to the end of the processing and the user terminal 59 returns to its idle state and awaits the next input query. If, however, the user indicates (by, for example, inputting an appropriate voice command) that the search results do not correspond to the desired information, then the processing proceeds from step s11 to step s13, where the search engine 53 performs a phoneme search of the database 29. However, in this embodiment, the phoneme search performed in step s13 is not of the whole database 29, since this could take several hours depending on the size of the database 29.

Instead, the phoneme search performed in step s13 uses the results of the word search performed in step s5 to

identify one or more portions within the database which may correspond to the user's input query. The way in which the phoneme search performed in step s13 is performed in this embodiment, will be described in more detail later. After the phoneme search has been performed, the control unit 55 identifies, in step s15, if a match has been found. If a match has been found, then the processing proceeds to step s17 where the control unit 55 causes the search results to be displayed to the user on the display 57. Again, the system then awaits the user's confirmation as to whether or not the search results correspond to the desired information. If the results are correct, then the processing passes from step s19 to the end and the user terminal 59 returns to its idle state and awaits the next input query. If however, the user indicates that the search results do not correspond to the desired information, then the processing proceeds from step s19 to step s21, where the control unit 55 is operable to ask the user, via the display 57, whether or not a phoneme search should be performed of the whole database 29. If in response to this query, the user indicates that such a search should be performed, then the processing proceeds to step s23 where the search engine performs a phoneme search of the entire database 29.

On completion of this search, the control unit 55

identifies, in step s25, whether or not a match for the user's input query has been found. If a match is found, then the processing proceeds to step s27 where the control unit 55 causes the search results to be displayed to the user on the display 57. If the search results are correct, then the processing proceeds from step s29 to the end of the processing and the user terminal 59 returns to its idle state and awaits the next input query. If, on the other hand, the user indicates that the search results still do not correspond to the desired information, then the processing passes to step s31 where the control unit 55 queries the user, via the display 57, whether or not the user wishes to redefine or amend the search query. If the user does wish to redefine or amend the search query, then the processing returns to step s3 where the user's subsequent input query is processed in a similar manner. If the search is not to be redefined or amended, then the search results and the user's initial input query are discarded and the user terminal 59 returns to its idle state and awaits the next input query.

PHONEME SEARCH

As mentioned above, in steps s13 and s23, the search engine 53 compares the phoneme data of the input query with the phoneme data in the phoneme and word lattice annotation data stored in the database 29. Various

techniques can be used including standard pattern matching techniques such as dynamic programming, to carry out this comparison. In this embodiment, a technique which we refer to as M-GRAMS is used. This technique was proposed by Ng, K. and Zue, V.W. and is discussed in, for example, the paper entitled "Subword unit representations for spoken document retrieval" published in the proceedings of Eurospeech 1997.

The problem with searching for individual phonemes is that there will be many occurrences of each phoneme within the database. Therefore, an individual phoneme on its own does not provide enough discriminability to be able to match the phoneme string of the input query with the phoneme strings within the database. Syllable sized units, however, are likely to provide more discriminability, although they are not easy to identify. The M-GRAM technique presents a suitable compromise between these two possibilities and takes overlapping fixed size fragments, or M-GRAMS, of the phoneme string to provide a set of features. This is illustrated in Figure 8, which shows part of an input phoneme string having phonemes a, b, c, d, e, and f, which are split into four M-GRAMS (a, b, c), (b, c, d), (c, d, e) and (d, e, f). In this illustration, each of the four M-GRAMS comprises a sequence of three phonemes which is unique and represents a unique feature (f_i) which can be found

within the input phoneme string.

Therefore, referring to Figure 7, the first step s51 in performing the phoneme search in step s13 shown in Figure 6, is to identify all the different M-GRAMS which are in the input phoneme data and their frequency of occurrence. Then, in step s53, the search engine 53 determines the frequency of occurrence of the identified M-GRAMS in the selected portion of the database (identified from the word search performed in step s5 in Figure 6). To illustrate this, for a given portion of the database and for the example M-GRAMS illustrated in Figure 8, this yields the following table of information:

M-GRAM (feature (f_i))	Input phoneme string frequency of occurrence ($\underline{q}$)	Phoneme string of selected portion of database ($\underline{a}$)
M_1	1	0
M_2	2	2
M_3	3	2
M_4	1	1

Next, in step s55, the search engine 53 calculates a similarity score representing a similarity between the phoneme string of the input query and the phoneme string of the selected portion from the database. In this embodiment, this similarity score is determined using a

cosine measure using the frequencies of occurrence of the identified M-GRAMS in the input query and in the selected portion of the database as vectors. The philosophy behind this technique is that if the input phoneme string is similar to the selected portion of the database phoneme string, then the frequency of occurrence of the M-GRAM features will be similar for the two phoneme strings. Therefore, if the frequencies of occurrence of the M-GRAMS are considered to be vectors (i.e. considering the second and third columns in the above table as vectors), then if there is a similarity between the input phoneme string and the selected portion of the database, then the angle between these vectors should be small. This is illustrated in Figure 9 for two-dimensional vectors $\underline{a}$ and $\underline{q}$, with the angle between the vectors given as θ. In the example shown in Figure 8, the vectors $\underline{a}$ and $\underline{q}$ will be four dimensional vectors and the similarity score can be calculated from:

$$SCORE = COS\,\theta = \frac{\underline{a} \cdot \underline{q}}{|\underline{a}|\,|\underline{q}|}$$

This score is then associated with the current selected portion of the database and stored until the end of the search. In some applications, the vectors used in the calculation of the cosine measure will be the logarithm of these frequencies of occurrences, rather than the

frequencies of occurrences themselves.

The processing then proceeds to step s57 where the search engine 53 identifies whether or not there are any more selected portions of phoneme strings from the database 29. If there are, then the processing returns to step s53 where a similar procedure is followed to identify the score for this portion of the database. If there are no more selected portions, then the searching ends and the processing returns to step s15 shown in Figure 6, where the control unit considers the scores generated by the search engine 53 and identifies whether or not there is a match by, for example, comparing the calculated scores with a predetermined threshold value.

As those skilled in the art will appreciate, a similar matching operation will be performed in step s23 shown in Figure 6. However, since the entire database is being searched, this search is carried out by searching each of the blocks discussed above in turn.

As those skilled in the art will appreciate, this type of phonetic and word annotation of input text document provides a convenient and powerful way to allow a user to later search the database by voice for the text document. Additionally, this way of annotating the documents results in a more robust annotation, since the

optical character recognition unit 105 may incorrectly recognise one or more characters within the text of the input document.

ALTERNATIVE EMBODIMENTS

In the above embodiments, a phonetic transcription unit 107 was used for generating the annotation data for annotating the image or text data. As those skilled in the art will appreciate, other techniques can be used. For example, a human operator can manually generate this annotation data from the image of the document itself.

In the above embodiment, the database 29 and the automatic speech recognition unit were both located within the user terminal 59. As those skilled in the art will appreciate, this is not essential. Figure 10 illustrates an embodiment in which the database 29 and the search engine 53 are located in a remote server 60 and in which the user terminal 59 accesses the database 29 via the network interface units 67 and 69 and a data network 68 (such as the internet). In operation, the user inputs a voice query via the microphone 7 which is converted into phoneme and word data by the automatic speech recognition unit 51. This data is then passed to the control unit which controls the transmission of this phoneme and word data over the data network 68 to the search engine 53 located within the remote server 60.

The search engine 53 then carries out the search in a similar manner to the way in which the search was performed in the first embodiment. The results of the search are then transmitted back from the search engine 53 to the control unit 55 via the data network 68. The control unit considers the search results received back from the network and displays appropriate data on the display 57 for viewing by the user 39.

In addition to locating the database 29 and the search engine 53 in the remote server 60, it is also possible to locate the automatic speech recognition unit 51 in the remote server 60. Such an embodiment is shown in Figure 11. As shown in this embodiment, the input voice query from the user is passed via input line 61 to a speech encoding unit 73 which is operable to encode the speech for efficient transfer through the data network 68. The encoded data is then passed to the control unit 55 which transmits the data over the network 68 to the remote server 60, where it is processed by the automatic speech recognition unit 51. The phoneme and word data generated by the speech recognition unit 51 for the input query is then passed to the search engine 53 for use in searching the database 29. The search results generated by the search engine 53 are then passed, via the network interface 69 and the network 68, back to the user terminal 59. The search results received back from the

remote server are passed via the network interface unit 67 to the control unit 55 which analyses the search results and generates and displays appropriate data on the display 57 for viewing by the user.

CLAIMS:

1. An apparatus for generating annotation data for use in annotating a data file, the apparatus comprising:

means for receiving image data representative of a text document;

character recognition means for converting said image data into text data; and

means for generating annotation data defining a phoneme and word lattice for the text data; and

wherein said generating means comprises:

(i) means for generating data defining a plurality of nodes within the lattice and a plurality of links connecting the nodes within the lattice; and

(ii) means for generating data associating each phoneme within the text data with a respective link within said lattice and for associating words within the text data with a respective link within said lattice.

2. An apparatus according to claim 1, wherein said generating means is operable to generate said data defining said phoneme and word lattice in blocks of said nodes.

3. An apparatus according to claim 1 or 2, wherein said

generating means is operable to generate data defining time stamp information for each of said nodes.

4. An apparatus according to claim 3, wherein said generating means is arranged to generate said phoneme and word lattice data in blocks of equal time duration.

5. An apparatus according to claim 2 or 4, wherein said generating means is operable to generate data which defines each block's location within a database.

6. An apparatus according to any preceding claim, wherein said means for defining a plurality of nodes and a plurality of links is operable to define at least one node which is connected to a plurality of other nodes by a plurality of links.

7. An apparatus according to claim 6, wherein at least one of said plurality of links connecting said nodes to said plurality of other nodes is associated with a phoneme and wherein at least one of said links connecting said node to said plurality of other nodes is associated with a word.

8. An apparatus according to any preceding claim, wherein said generating means comprises an automatic phonetic transcription unit which generates said phoneme

and word lattice annotation data in response to the text data output by said converter.

9. An apparatus according to any preceding claim, further comprising means for associating said annotation data with either said image data representative of said text document or with said text data.

10. An apparatus according to any preceding claim, wherein said receiving means comprises a document scanner or a facsimile machine.

11. A method of generating annotation data for use in annotating a data file, the method comprising the steps of:

receiving image data representative of a text document;

converting said image data into text data using a character recognition unit; and

generating annotation data defining a phoneme and word lattice for the input voice annotation signal;

wherein said generating step comprises the steps of:

(i) generating data defining a plurality of nodes within the lattice and a plurality of links connecting the nodes within the lattice; and

(ii) generating data associating each phoneme within said text data with a respective link

within said lattice and associating each word within said text data with a respective link within said lattice.

12. A method according to claim 11, wherein said generating step generates said data defining said phoneme and word lattice in blocks of said nodes.

13. A method according to claim 11 or 12, wherein said generating step generates data defining time stamp information for each of said nodes.

14. A method according to claim 13, wherein said generating step generates said phoneme and word lattice data in blocks of equal time duration.

15. A method according to claim 12 or 13, wherein said generating step generates data which defines each block's location within a database.

16. A method according to any of claims 11 to 15, wherein said step of defining a plurality of nodes and a plurality of links defines at least one node which is connected to a plurality of other nodes by a plurality of links.

17. A method according to claim 16, wherein at least one

of said plurality of links connecting said node to said plurality of other nodes is associated with a phoneme and wherein at least one of said links connecting said node to said plurality of other nodes is associated with a word.

18. A method according to any of claims 11 to 17, wherein said generating step uses an automatic phonetic transcription unit which generates said phoneme and word lattice annotation data in response to the text data output by said converting step.

19. A method according to any of claims 11 to 18, further comprising the step of associating said annotation data with either said received image data or with said text data.

20. A method according to any preceding claim, wherein said receiving step uses a document scanner or a facsimile machine.

ABSTRACT

DATABASE ANNOTATION AND RETRIEVAL

A data structure is provided for annotating data files within a database. The annotation data comprises a phoneme and word lattice which allows the quick and efficient searching of data files within the database, in response to a user's input query for desired information. The structure of the annotation data is such that it allows the input query to be made by voice and can be used for annotating various kinds of data files, such as audio data files, audio and visual data files, multimedia data files etc.

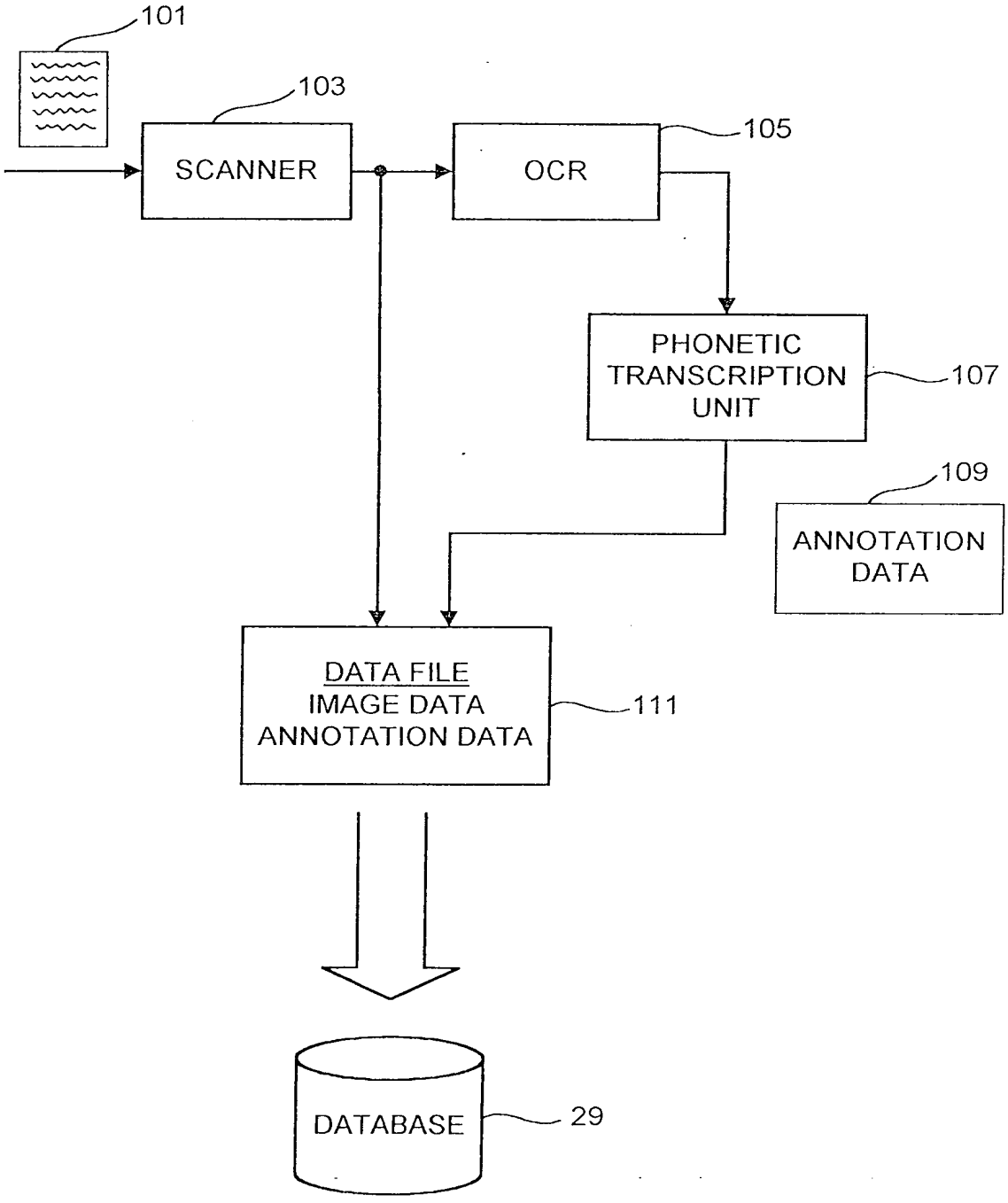
101
103
SCANNER
105
OCR
PHONETIC
TRANSCRIPTION
UNIT
107
109
ANNOTATION
DATA
DATA FILE
IMAGE DATA
ANNOTATION DATA
111
DATABASE
29

Fig. 1

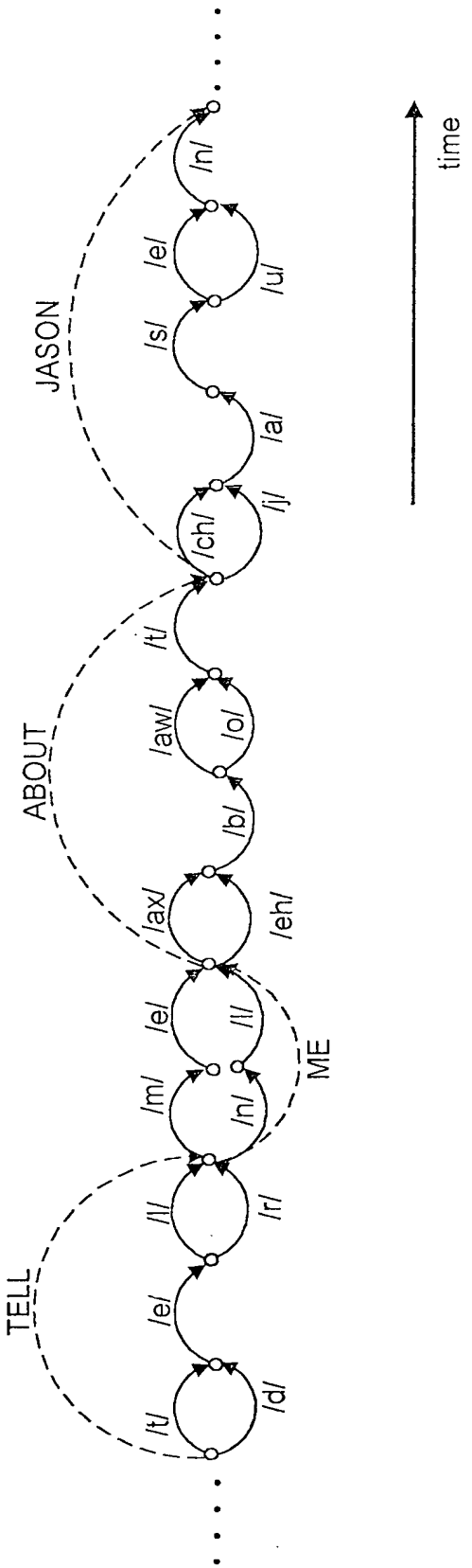

Fig. 2

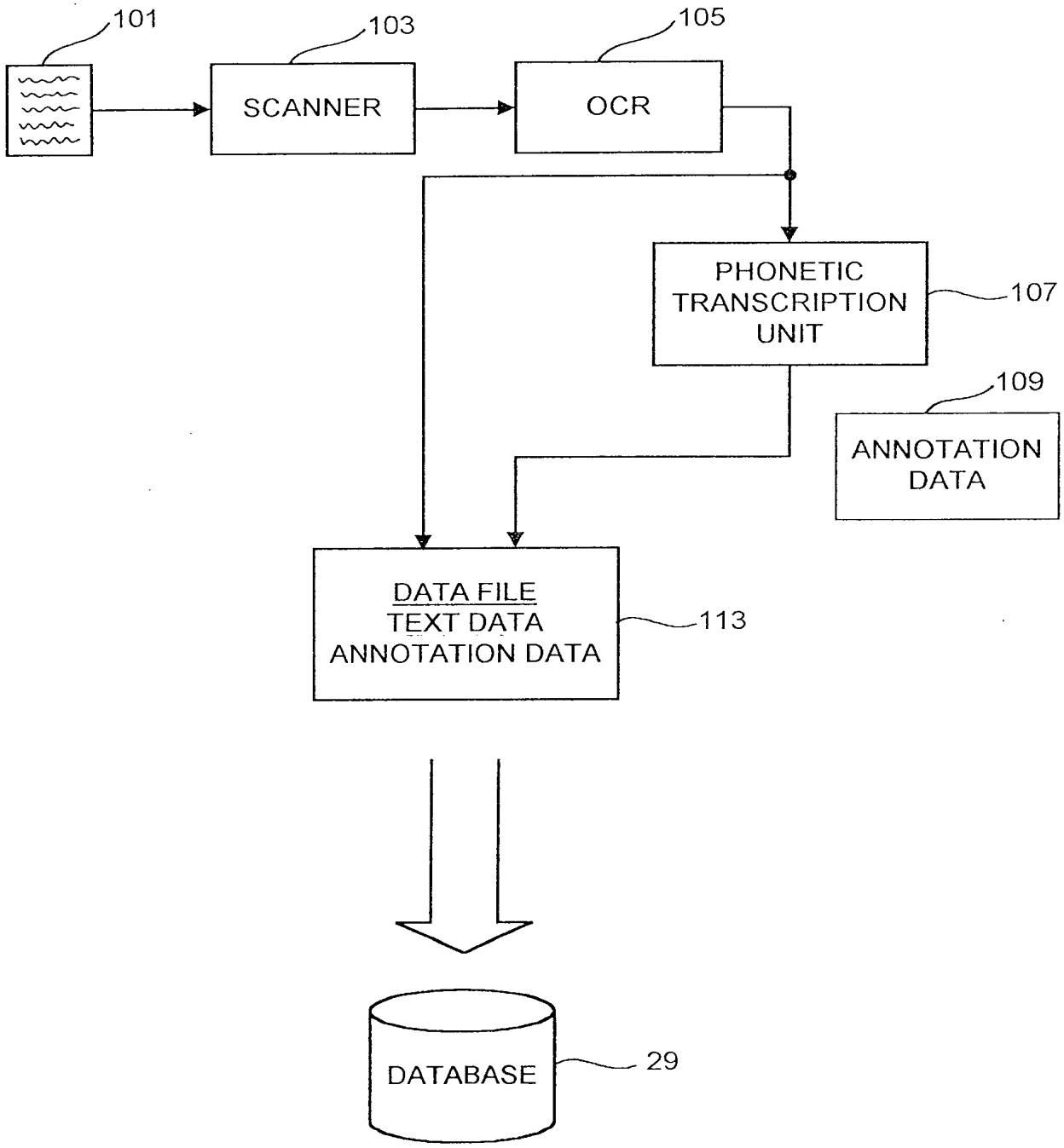
101
103
105
SCANNER
OCR
PHONETIC TRANSCRIPTION UNIT
107
109
ANNOTATION DATA
DATA FILE
TEXT DATA
ANNOTATION DATA
113
DATABASE
29

Fig. 3

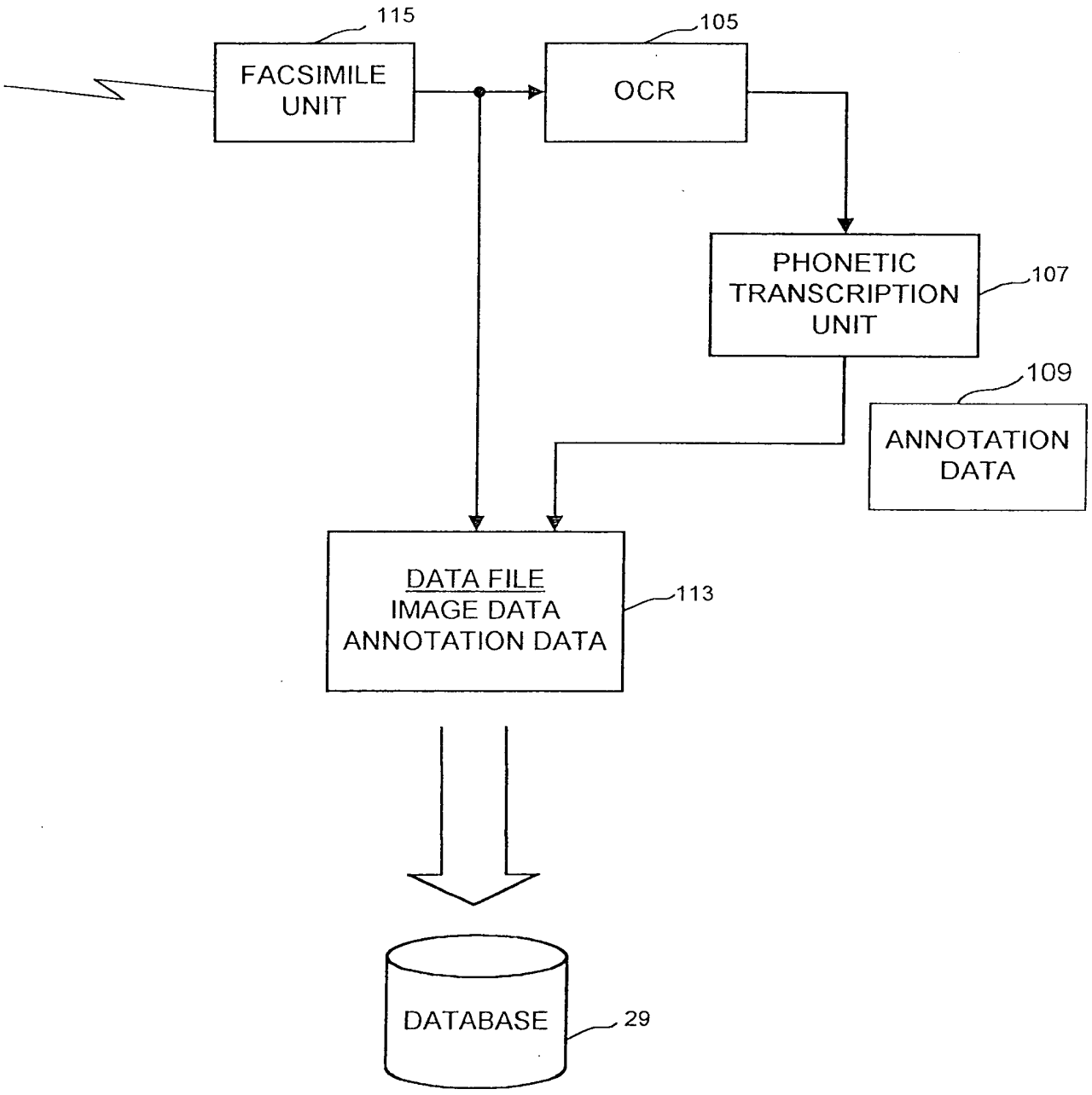
115
FACSIMILE UNIT
105
OCR
PHONETIC TRANSCRIPTION UNIT
107
109
ANNOTATION DATA
DATA FILE
IMAGE DATA
ANNOTATION DATA
113
DATABASE
29

Fig. 4

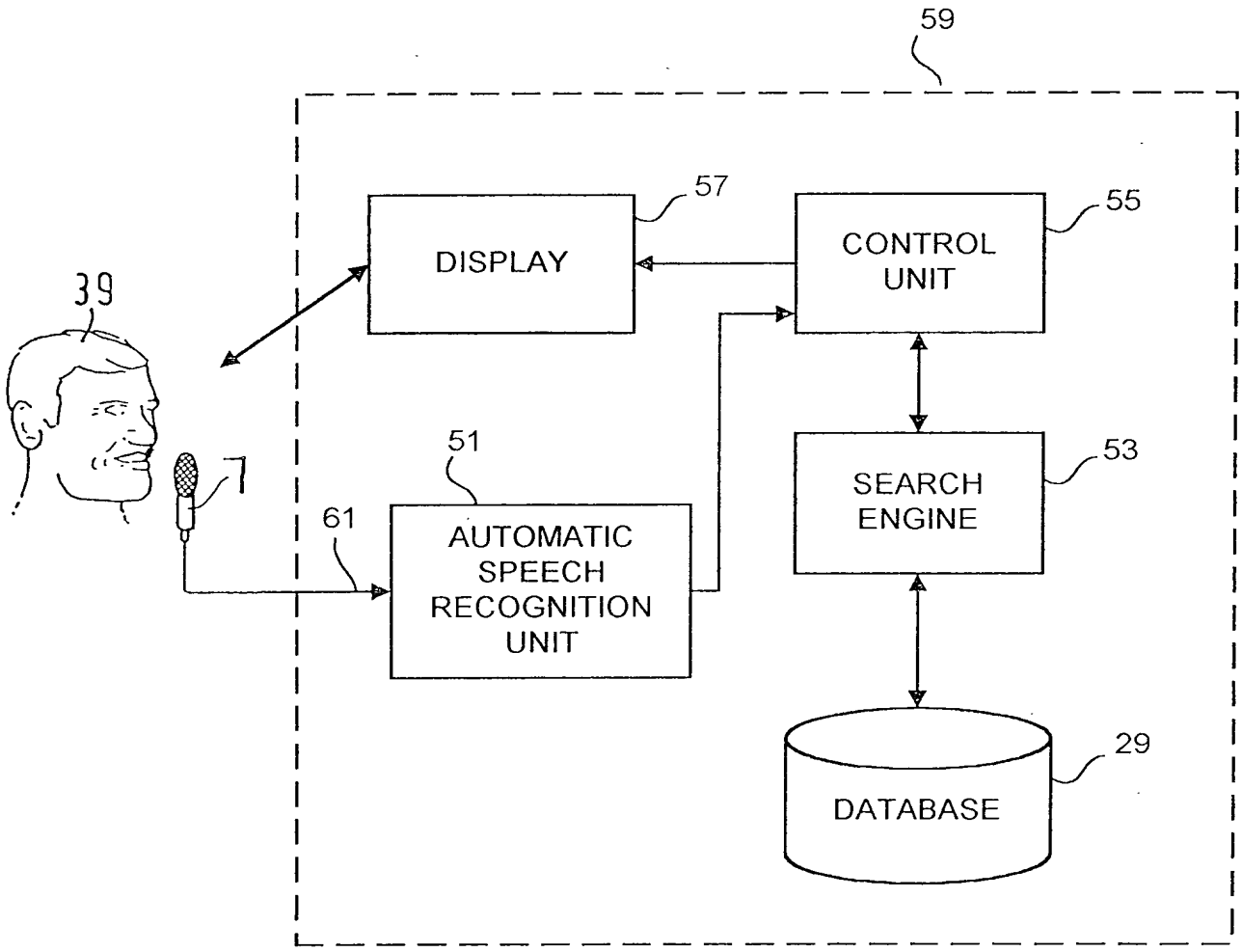
59
57
DISPLAY
55
CONTROL UNIT
39
7
51
53
SEARCH ENGINE
61
AUTOMATIC SPEECH RECOGNITION UNIT
29
DATABASE

Fig. 5

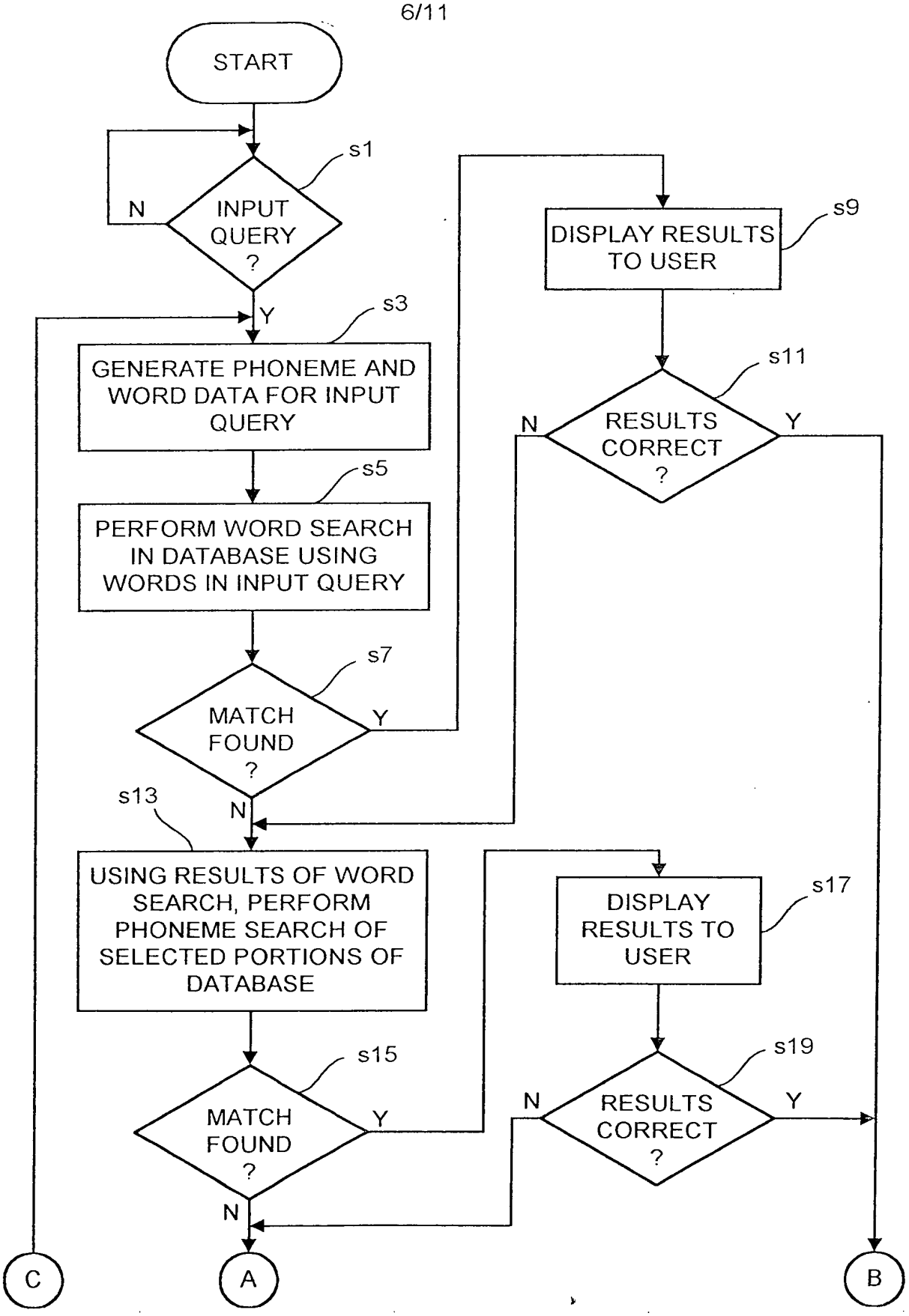
START
s1
INPUT QUERY ?
N
Y
s3
GENERATE PHONEME AND WORD DATA FOR INPUT QUERY
s5
PERFORM WORD SEARCH IN DATABASE USING WORDS IN INPUT QUERY
s7
MATCH FOUND ?
Y
N
s13
USING RESULTS OF WORD SEARCH, PERFORM PHONEME SEARCH OF SELECTED PORTIONS OF DATABASE
s15
MATCH FOUND ?
Y
N
C
A
s9
DISPLAY RESULTS TO USER
s11
RESULTS CORRECT ?
N
Y
s17
DISPLAY RESULTS TO USER
s19
RESULTS CORRECT ?
N
Y
B

Fig. 6a

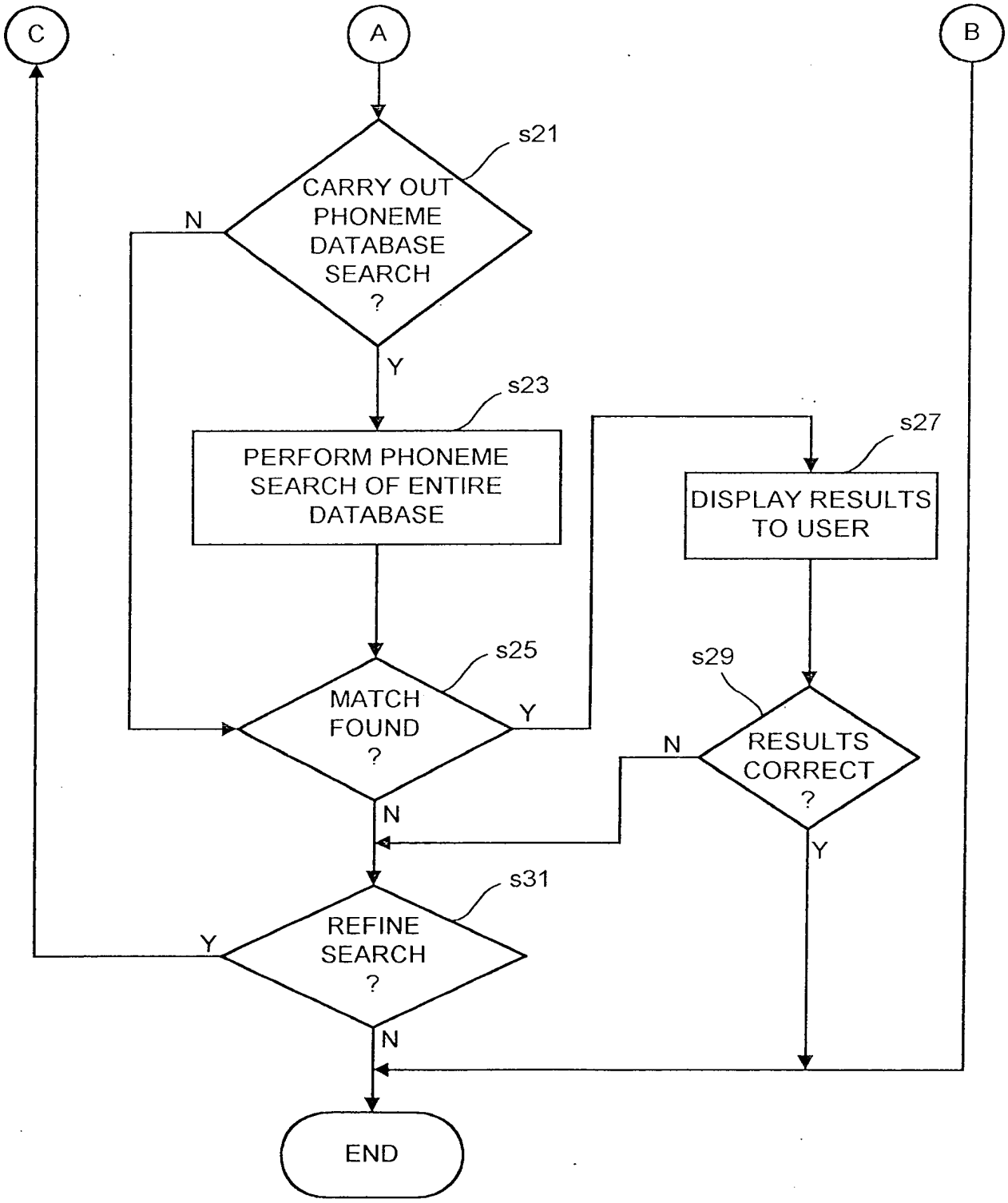
C
A
B
s21
CARRY OUT PHONEME DATABASE SEARCH ?
N
Y
s23
PERFORM PHONEME SEARCH OF ENTIRE DATABASE
s27
DISPLAY RESULTS TO USER
s25
MATCH FOUND ?
Y
N
s29
RESULTS CORRECT ?
N
Y
s31
REFINE SEARCH ?
Y
N
END

Fig. 6b

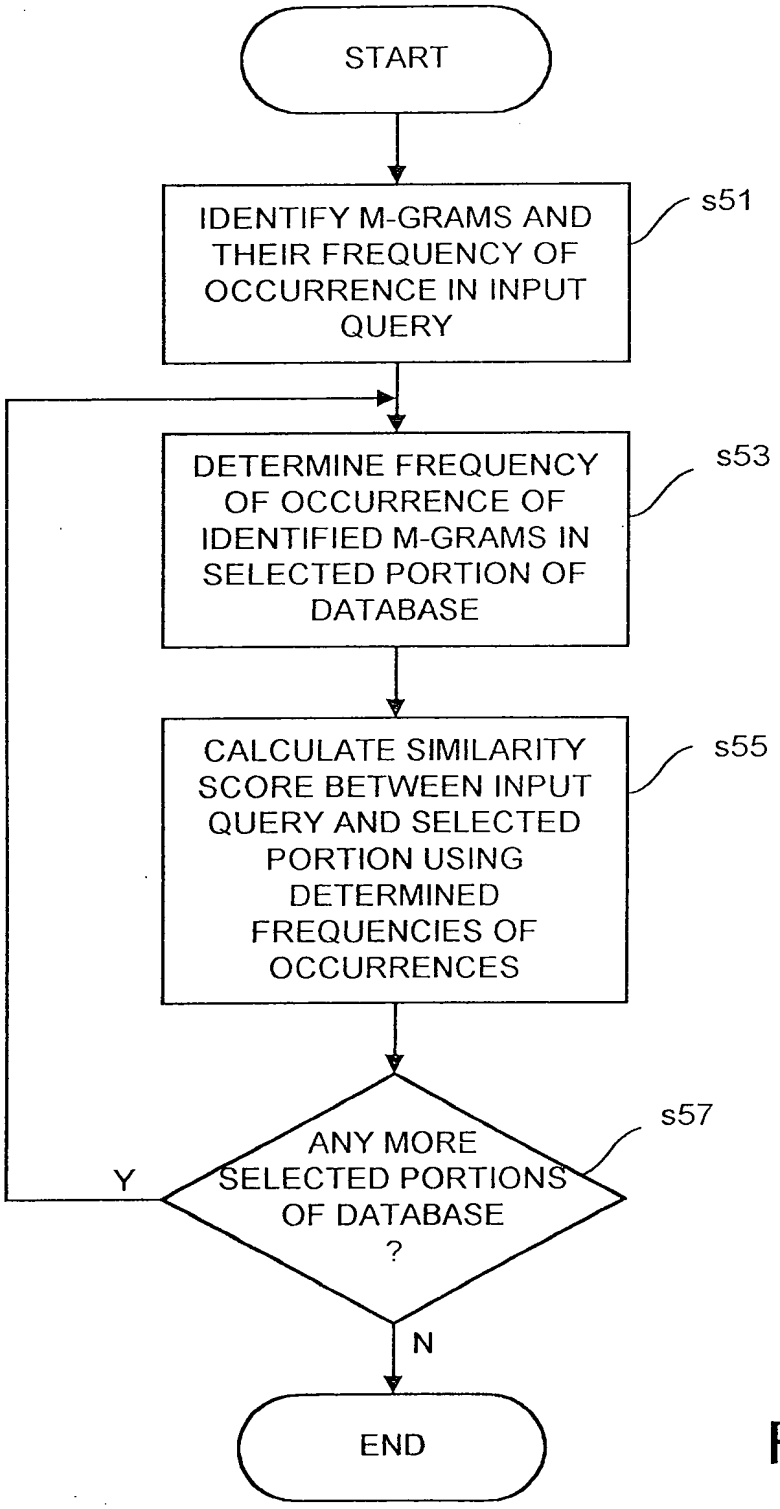
START
IDENTIFY M-GRAMS AND THEIR FREQUENCY OF OCCURRENCE IN INPUT QUERY
s51
DETERMINE FREQUENCY OF OCCURRENCE OF IDENTIFIED M-GRAMS IN SELECTED PORTION OF DATABASE
s53
CALCULATE SIMILARITY SCORE BETWEEN INPUT QUERY AND SELECTED PORTION USING DETERMINED FREQUENCIES OF OCCURRENCES
s55
ANY MORE SELECTED PORTIONS OF DATABASE ?
s57
Y
N
END

Fig.7

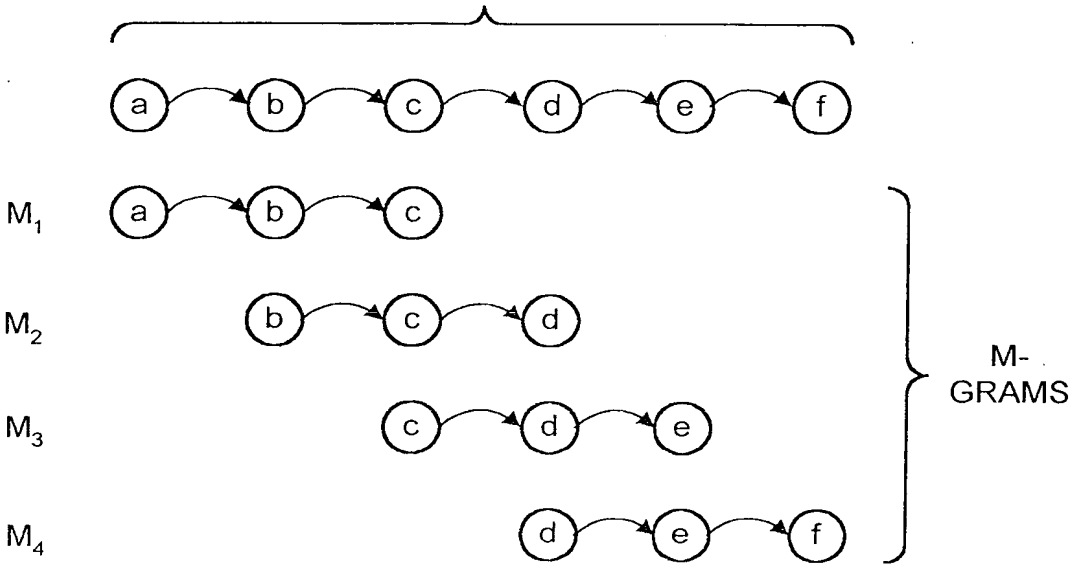

Fig. 8

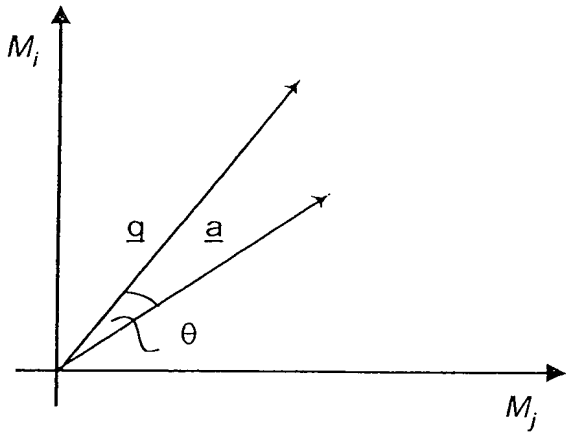

Fig. 9

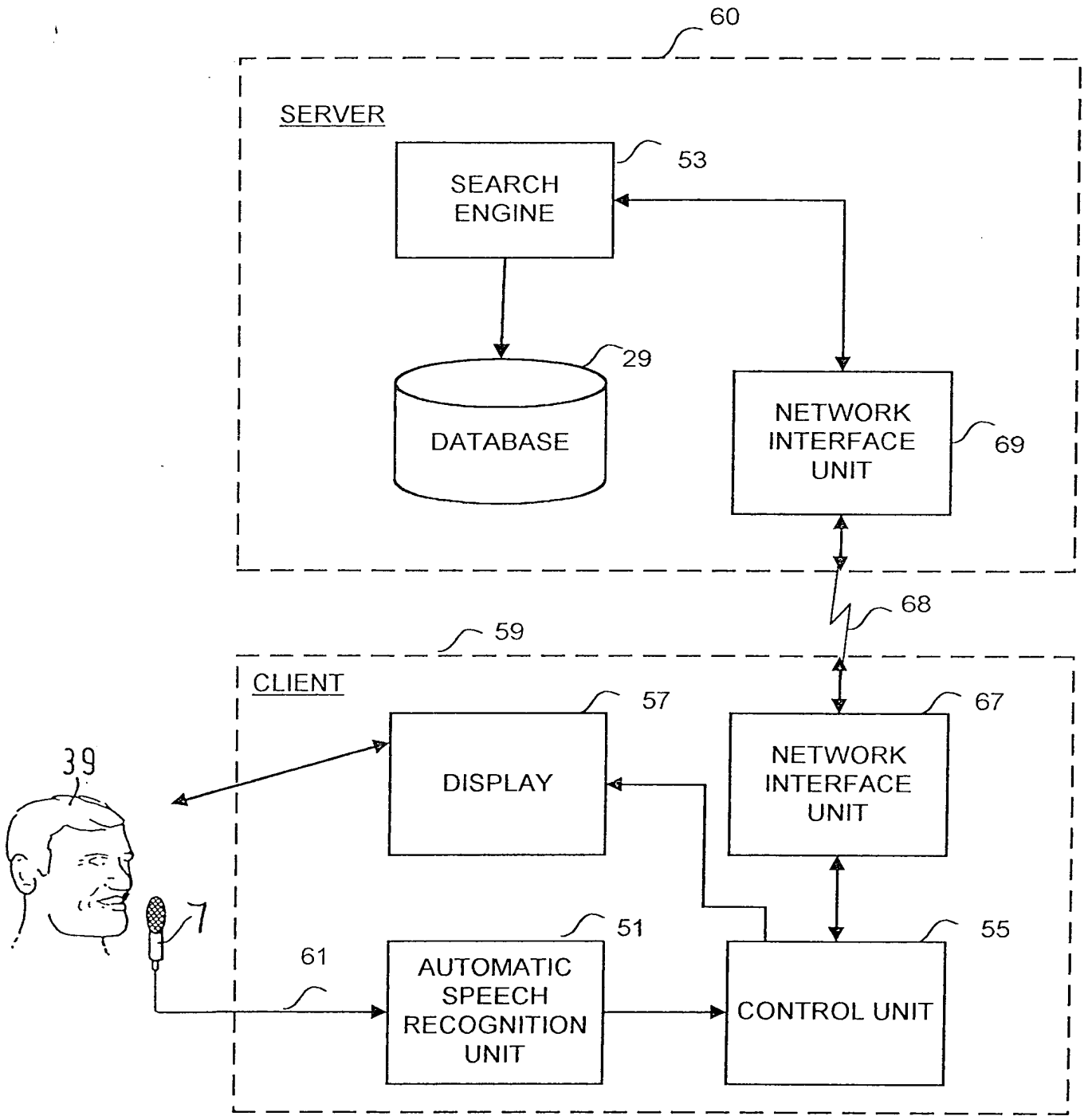
60
SERVER
SEARCH ENGINE
53
29
DATABASE
NETWORK INTERFACE UNIT
69
68
59
CLIENT
57
DISPLAY
67
NETWORK INTERFACE UNIT
39
7
61
51
AUTOMATIC SPEECH RECOGNITION UNIT
55
CONTROL UNIT

Fig. 10

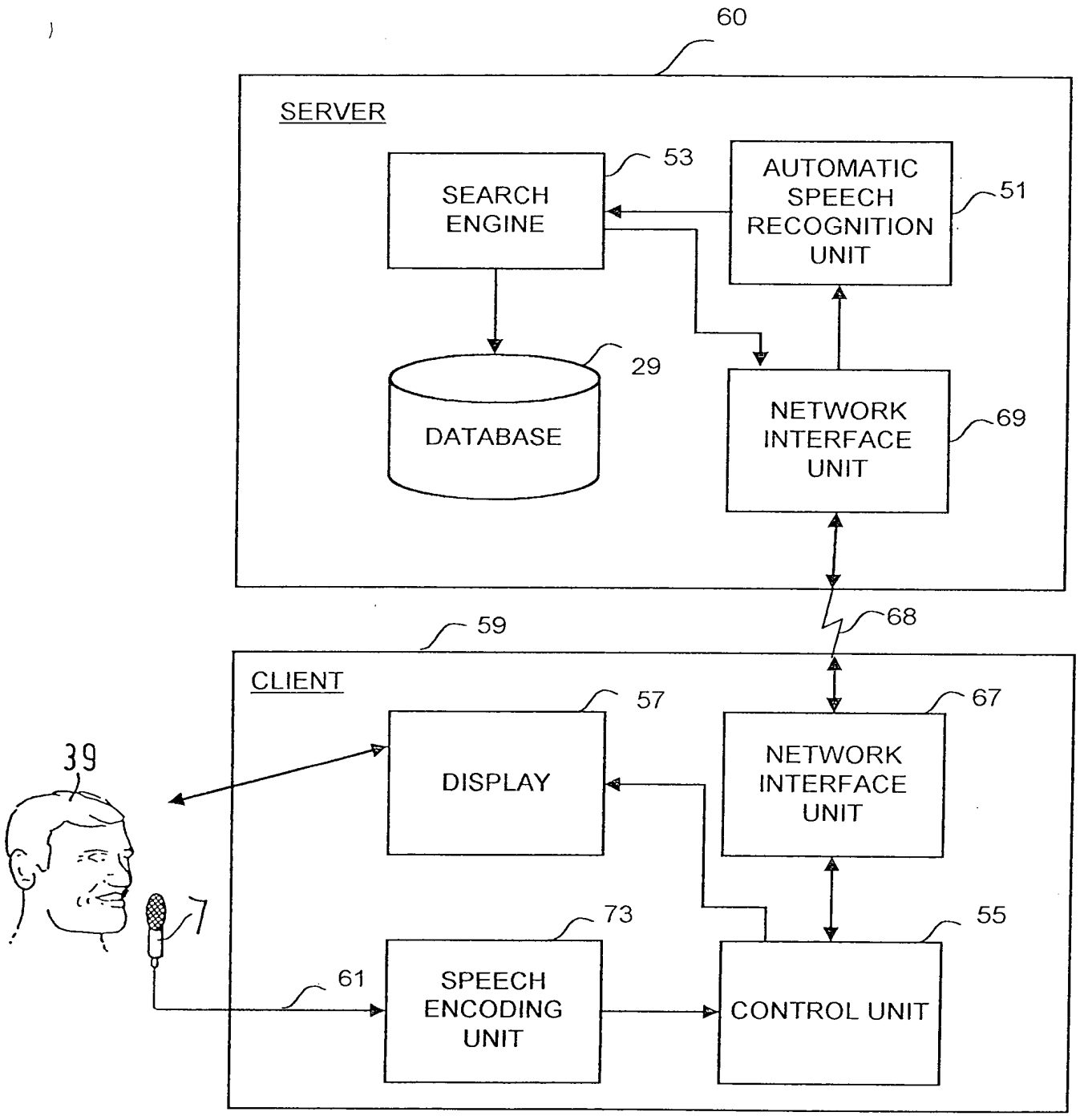
60
SERVER
SEARCH ENGINE
53
AUTOMATIC SPEECH RECOGNITION UNIT
51
DATABASE
29
NETWORK INTERFACE UNIT
69
68
59
CLIENT
57
DISPLAY
67
NETWORK INTERFACE UNIT
39
7
73
55
61
SPEECH ENCODING UNIT
CONTROL UNIT

Fig. 11